Beyond age comparisons: a plea for the use of a modified Brunswikian approach to experimental designs in the study of adult development and aging

Freund, Alexandra M ; Isaacowitz, Derek M

DOI: https://doi.org/10.1159/000357177

Originally published at:
Freund, Alexandra M; Isaacowitz, Derek M (2013). Beyond age comparisons: a plea for the use of a modified Brunswikian approach to experimental designs in the study of adult development and aging. Human Development, 56(6):351-371.
DOI: https://doi.org/10.1159/000357177

Original Paper

Human Development 2013;56:351–371
DOI: 10.1159/000357177

Human Development

Beyond Age Comparisons: A Plea for the Use of a Modified Brunswikian Approach to Experimental Designs in the Study of Adult Development and Aging

Alexandra M. Freund[a] Derek M. Isaacowitz[b]

[a]University of Zurich, Zurich, Switzerland; [b]Northeastern University, Boston, Mass., USA

Key Words
Adult development and aging · Age group comparisons · Brunswik · Experimental design · Methods

Abstract
This article makes a plea for experimental designs in the field of adult development and aging using an approach to research inspired by the work of Egon Brunswik. Our recommendations are intended to complement correlational approaches and to enhance the testing of explanatory mechanisms. Our arguments are predicated on the fact that the field of adult development and aging faces particular methodological challenges stemming from the investigation of individual differences approached with age group comparison designs. Many studies on adult development and aging use extreme-group comparisons, contrasting young and older adults, although such comparisons can lead to the overestimation of age-related effects. Moreover, age group membership is used as a proxy variable for psychological processes leading to the observed age-related differences. The inherent correlational design of such age group comparisons can only approximate a test of the underlying psychological processes causing the differences between the groups. We consider these problems and potential solutions to them involving a Brunswikian approach to experimental design in research on adult development and aging, and we discuss implications for theory-predicated research in other subfields of developmental science with similar methodological issues.

Provocatively, Birren [1999] stated about 15 years ago that aging research is "data rich and theory poor" (p. 459). In this article, we add to this assessment that the field of adult development and aging is still theoretically underdeveloped because of the specific methodological challenges plaguing this field. We advocate the use of an approach to experimental designs in the study of adult development and aging that is inspired by the work of Egon Brunswik [1952, 1955], and we will argue that re-

KARGER

E-Mail karger@karger.com
www.karger.com/hde

0018–716X/14/0566–0351$39.50/0

Alexandra M. Freund
University of Zurich, Department of Psychology
Binzmühlestrasse 14, PO Box 11
CH–8050 Zurich (Switzerland)
E-Mail freund@psychologie.uzh.ch

searchers in this area need to specify their theoretical assumptions about the changing interrelationships between persons and their environment across adulthood. In turn, supplementing the Brunswikian approach with experimental designs manipulating the hypothesized processes leading to developmental change can contribute significantly to theories of adult development and aging.

Such an approach is also in line with relational developmental systems models [e.g., Lerner, 2012; Overton, 2013] that stress the multidirectional transaction between a person and his/her context. The relational developmental systems approach posits that developmental processes can only be understood in terms of adaptations to a person's ecology, which, in turn, is reciprocally shaped by the person. This relational model of development goes far beyond the analysis of context by person-variable interactions, and instead focuses on the mutually constitutive relations of all levels of developmental systems spanning cells to societies. In our view, this relational approach calls for specifying the processes of adapting to an ecology depending on the developmental status of the person and, as well, the changes in the environment that a person brings about by interacting with it. Similarly, at the heart of Brunswik's approach lies the assumption that psychological processes are adaptations to the ecology of a person. Importantly, Brunswik posits that psychological processes can be tested adequately only if the stimuli are representative of the ecology of the person under investigation. To the extent that theoretical models posit changes in psychological adaptations with age, empirical approaches need to directly investigate potential ecological changes as well.

Recently, psychology in general, and social and experimental psychology in particular, has seen a lively upsurge of the debate on adequate methods to reveal reliable and valid results. For instance, Fiedler [2011] pointed out that the practice of how experimental psychologists select stimulus materials, experimental manipulations, and the assessment of dependent variables contribute to the maximization of effect sizes. Similarly, Simmons, Nelson, and Simonsohn [2011] have stimulated a debate concerning how certain research practices regarding sample size or the collapsing of different conditions might lead to a high false-positive rate regarding the presence of effects. Although we agree with many of the important issues raised by Fiedler [2011] as well as Simmons et al. [2011], we do not want to reiterate the arguments and recommendations here. Instead, we want to focus on the specific methodological challenges facing research investigating individual differences using group comparison designs such as is commonly done in the area of adult development and aging. Interestingly, this field may be somewhat less susceptible to several of the specific issues raised in the debate on false positivity, given how enormously resource intensive it is to collect data on adults of different ages. The large effort associated with data collection might lead researchers to be more willing to report null results and less willing to throw out entire conditions of studies. However, the very resource intensiveness of research with non-student populations might also prevent researchers from attempting replications because they tend to be less likely to be published in top journals and are often seen as merely incremental.

In this article, we focus on research on adult development and aging, but maintain that the challenges and the possible solutions may hold for other psychological subdisciplines with similar methodological issues such as personality, clinical, or cultural psychology. These issues pertain particularly to the reliance on extreme-group comparisons and to the role of analytic methods for deriving and testing hypotheses. We also consider several concerns more specific to the field of adult development and aging, including the use of cross-sectional designs as well as the common use of mediation analyses in the pursuit of causal factors underlying age-related differences [for this latter point, see Lindenberger, von Oertzen, Ghisletta, & Hertzog, 2011].

After considering these issues, we present several recommendations for moving these fields ahead in ways that are methodologically sound as well as conceptually appropriate for investigating age-related differences. We present problems first, then possible recommendations separately, because the recommendations are not specific solutions to the problems raised; rather, we see them as best practices given the constraints inherent in behavioral research. More specifically, we stress the importance of testing theoretically derived hypotheses about developmental processes underlying differences between age groups, necessitating that researchers complement correlational studies with experiments to test the developmental mechanisms driving age-related differences.

We acknowledge that experimental manipulation of candidate mechanisms causing age-related differences is not always possible, for instance when studying long-term cognitive effects of health-related decline. In many cases, however, experiments are possible, for instance when studying effects of future time perspective on social selectivity [Fredrickson & Carstensen, 1990], the effect of limited cognitive resources on social information processing [Hess, Rosenberg, & Waters, 2001], or the effect of sensory functioning on cognitive performance [Lindenberger, Scherer, & Baltes, 2001].

Moreover, to avoid the potential pitfalls of experimentation related to the selection of the stimuli, the manipulation, and the dependent variables when using different age groups, we make a plea for complementing experiments designed to test specific developmental mechanisms with a Brunswikian approach [e.g., Brunswik, 1952, 1955]. Brunswik was a functionalist, whose central question was how organisms function in a given environment. Importantly, Brunswik placed equal emphasis on the organism and the environment and posited that advances in psychology hinge upon the conceptualization of the interrelationships between organism and environment. This perspective does not only stress the importance of theoretical specifications of the nature of the organism-environment interrelation but also has far-reaching consequences for the empirical investigations and proper experimentation, as will be elaborated in more detail in the last section of this paper. Currently, such an experimental approach is underutilized, and we hope to push researchers to expand their portfolio in order to incorporate such designs. Note that we do not claim that the study of adult development and aging does not include a large number of experiments. It does, particularly in the area of cognitive development. However, these experiments are typically still correlational, in that they compare age groups (e.g., a comparison of task-set switching costs of younger and older adults) and mostly do not manipulate experimentally the *mechanism* that might contribute to the age differences. Although such a manipulation of potential developmental mechanisms in the laboratory may not represent a specifically Brunswikian approach, we believe that a Brunswikian approach combined with such direct manipulation will ultimately be a very fruitful combination of approaches for the field.

Challenges to Research on Age-Related Differences

The study of adult development and aging faces unique challenges distinct from child development or other research on individual differences. A key challenge is the mundane fact that development during adulthood and aging occurs on a fairly long

DOI: 10.1159/000357177

time scale of 20, 30, 40, 50 or more years, rendering longitudinal studies extremely time-consuming and presenting a formidable challenge to the organization of a laboratory (e.g., keeping track of participants over decades and being able to match them over time without compromising data security). Importantly, researchers would need to wait for several decades before finishing data collection, leave alone publication of the data. By this time, psychology as a field will have moved on and is likely to be no longer interested in the research questions of the original study. These are purely pragmatic reasons that render longitudinal designs in this field difficult.

There are also methodological reasons that have been elaborated in depth and discussed at length in the life-span literature, so we do not need to elaborate on them here. Instead, we note the two most serious ones in passing: Selective sample attrition, and the confounding of age-graded, cohort-graded, and normative history-graded influences on development with changes over time. Baltes [1968] proposed to combine cross-sectional and longitudinal designs with multiple cohorts [see also Baltes, Reese, & Nesselroade, 1977; Schaie & Baltes, 1975], allowing the investigation of intraindividual change, interindividual differences in change as well as differences between cohorts [for an excellent recent use of this design, see Gerstorf, Ram, Hoppmann, Willis, & Schaie, 2011].

However, even such very costly designs (in terms of time and money) face the problem of selecting appropriate time intervals for the assessments. Lerner, Schwartz, and Phelps [2009] elaborated that although all developmental change can be plotted against the time scale of months, years, or decades, this does not imply that the months, years, or decades have the same developmental meaning for different developmental processes or variables. For instance, rapid and discontinuous change might occur over the course of only months in some variables during adolescence, but show a much slower and continuous rate of change over the years thereafter. During the same time intervals within the same people, other variables might show very different trajectories. The age-dependent differences in the trajectory of one variable and the differences between trajectories of different variables might be missed if using the same time intervals over the course of the longitudinal study. Therefore, Lerner et al. [2009] demand a theory-driven selection of the spacing of assessments.

Another problem concerns causality. Although longitudinal studies provide temporal information and thus allow so-called Granger causality statements (i.e., to test if one variable is able to forecast another), even the most complex longitudinal design is still correlational in nature, and thus has limitations in its ability to address *fully causal* factors contributing to developmental changes. Although comparisons of time-lagged associations between two (or more) variables can provide some evidence for which variable temporally precedes another one and, hence, is more likely to be the cause, the third-variable problem cannot be ruled out (i.e., that a third variable influences the temporal association of the two variables under consideration).

Propensity score methods are increasingly used in longitudinal designs in order to get a more nuanced approach to examining factors contributing to and processes underlying behavior in the context of correlational data [e.g., Jackson, Thoemmes, Jonkmann, Lüdtke, & Trautwein, 2012; Stuart, 2010]. In general, modeling dynamic change over time using extremely sophisticated multivariate statistical methods [e.g., McArdle, 2008] has made such impressive progress that the understandable excitement about these complex methods in the area of adult development and aging might overshadow the fact that no amount of sophistication in these methods can substitute

DOI: 10.1159/000357177

for experimental designs. Even with the most sophisticated methods for analyzing correlational data, the only way to address the third-variable problem and establish strict causality is experimental manipulation of the hypothesized causal variable, a point to which we return in more detail below.

We note, however, that noncausal correlations of some variables with age may be very important for practical purposes, although they may not be very satisfying from a theoretical perspective. As one practical example, the finding that vigilance for peripheral cues declines across adulthood may be quite important if the goal is to select the person best suited for the job of an air traffic controller, even if the available evidence is purely correlational. In this type of case, documenting the age-related decline in performance might be more important than delineating the underlying causes of the age-related decline. It is crucial for both the applied and the theoretical conclusions drawn from empirical findings that they result from the most appropriate design to study adult development. For instance, overestimating effect sizes of age-related decline might lead to unwarranted negative consequences for older adults (e.g., concerning chances on the job market). On the level of theory, inappropriate methodology could lead to faulty conclusions such as attesting lower emotional reactivity to older compared with younger adults when this result might be due to the selection of the specific emotional stimuli used in the study [cf. Kunzmann and Grühn, 2005; we will return to this example further below].

Problems Associated with Age Group Comparisons

Given the practical difficulties and interpretational problems with longitudinal designs, many studies in the field of adult development and aging (1) use cross-sectional age comparisons and interpret age-related differences as a proxy for developmental changes, (2) compare extreme groups of young and older adults and interpret differences between these two age groups as a monotonic function or even linear developmental change, and (3) use mediation analyses to test whether a potentially causal variable might mediate the relationship between age (as an independent variable) and some dependent variables of interest. Below, we review the key problems with each of these common practices.

Cross-Sectional Age Comparisons

Students and researchers in the field of adult development and aging are typically aware that cross-sectional comparison and extreme-group comparisons are highly flawed. Cross-sectional age comparisons are based on the assumption that differences *between* persons (young vs. old) are indicative of developmental change *within* persons. However, cross-sectional age-related differences confound age, cohort, and historical time [Baltes et al., 1977; Schaie, 1965]. Moreover, there might be multiple pathways to the same developmental outcomes or, in other words, interindividual differences in the factors causing developmental change. Discussion sections typically pay lip service to these limitations associated with cross-sectional designs; yet simply mentioning these problems does not remedy them and these designs nonetheless may provide flawed estimates of age-related differences.

Comparison of Extreme Age Groups

There are various methodological problems associated with extreme-group comparisons in all fields of psychology [for an excellent review, see Preacher, Rucker, MacCallum, & Nicewander, 2005]. Preacher et al. [2005] elaborate particularly on the issue that extreme-group comparisons lead to increased power and the increase in power is a direct function of the extremity of the groups. Similarly, extreme-group comparisons lead to inflated standardized effect sizes. Although some researchers might see this as an advantage as it increases the likelihood of obtaining significant and strong effects, it also increases the danger of producing false-positive results and of overestimating differences between the groups. Thus, although extreme-group comparisons might be cost effective in an individual study, they might be more cost intensive across laboratories and studies by overestimating effect sizes and producing false-positive results, which are then (erroneously) used as the basis for further studies. Moreover, as Preacher et al. [2005] put it: "The primary goal of research should not be to obtain significant p values but rather to determine what the data tell us about the phenomena of interest – that is, effect size and practical significance" (p. 181). Obviously, inflated effect sizes paint a distorted picture of the phenomena of interest and, thereby, might also lead to erroneous theoretical conclusions. To the extent that individual difference studies are tilted in favor of group difference studies in the first place, this orientation can create serious interpretive problems, to the extent that samples are selected that may inflate the extent of the actual differences in the population. More often than not, differences found for extreme groups are interpreted as reflecting the size of age-related differences in the general aging population.

Regarding age group comparisons, the first problem concerns the *selection of appropriate age groups*. To our knowledge, there are no theoretical models in the field of adult development and aging that delineate at which age exactly a specific developmental process manifests. This absence is due to the fact that there is much heterogeneity in development after childhood [e.g., Dannefer, 2003], although heterogeneity may again decline close to death [e.g., Wilson, Beck, Bienias, & Bennett, 2007]. According to Baltes, Reese, and Lipsitt [1980] this greater heterogeneity compared with child development might result from a stronger influence of nonnormative and, at least during young adulthood, also history-graded factors on development. Even such biological markers as menopause for entering middle adulthood in women exhibit large interindividual variability regarding the age at onset. Similarly, in many countries the age at which people transition into retirement, a marker that could delimit middle from old adulthood, can span a decade or more. Dannefer [2003] argues that cumulative advantages or disadvantages lead to interindividual divergence and, thereby, to increased age-related variability. In other words, developmental trajectories seem to become more varied with increasing age, making it difficult to provide exact age delimitations for developmental phenomena in adulthood and old age. This variation might also be the reason for the fact that there is currently no overarching developmental theory that provides age criteria for dividing adulthood into meaningful age groups. Together, this situation leads to ill-defined delimitations of young, middle, and older adulthood regarding chronological age range [for a discussion of the specific problem of middle adulthood, see Freund & Ritter, 2009]. This situation can lead to the problem that different studies use different age ranges for the same age

DOI: 10.1159/000357177

groups and may, as a consequence, come to different conclusions regarding the existence or magnitude of age-related differences.

One of the best solutions to this problem might be to forgo the formation of discrete age groups and use age as a continuous variable. This use would not only solve the problem of ill-defined age groups but also allow researchers to take full advantage of the variability associated with age. Unfortunately, this step is only rarely made, likely because it is very difficult to get the busy group of middle-aged adults who have to juggle the many demands of work and family to spend time participating in a study. Instead, a more pragmatic solution to the problem of ill-defined age groups that many studies follow is to recruit younger and older adults on the basis of who is easily accessible and willing to participate. Mostly, this procedure involves the recruitment of undergraduate students at the university campus of the respective research laboratory for the group of younger adults, and of older adults who either respond to advertisements in local newspapers or who are recruited at senior centers.

This recruitment practice can lead to various differences between the age groups that might be related to differences in age or to other differences associated with these two age groups (e.g., between students and nonstudents, or between people sharing rooms in dorms on a university campus and people living in a private household). As age cannot be randomly assigned, it is impossible to determine whether age-related differences are due to the differences between these groups in various background variables or whether they represent true age-related differences. Statistically controlling for age-related differences in such background variables, for instance as a covariate in an ANCOVA or by partialling it out in a regression analysis, may not remedy the problem. This ineffectiveness is due to the fact that the main effect of age after partialling out the variance of the control variable cannot be distinguished from the interaction effect of age on the control variable [Miller & Chapman, 2001]. In fact, the use of a covariate in the often-used ANCOVA in an age-comparative design is only justified if the distribution of the covariate is not associated with age [Elashoff, 1969]. According to Miller and Chapman [2001], "there is no statistical method that can address the question of whether two groups that differ on variable A would differ on variable B if they did not differ on variable A" (p. 43).

In the literature on aging, an often-used comparison group to older adults is the group of undergraduate students, who are typically quite similar to each other in terms of lifestyle, educational background, and their current occupation. By comparison, the group of older adults recruited to participate in psychological research is likely to have a much more diverse educational background, different occupational histories as well as lifestyles. Thus, when comparing younger with older adults, many studies compare a relatively homogeneous group of undergraduate students with a relatively diverse group of older adults. Moreover, the two groups often differ regarding the breadth of their age span. The group of younger adults often comprises the undergraduate years with an age span of about 7 years (mostly between 18 and 25 years). In contrast, the criterion for inclusion in the group as older adults is often that participants are older than 60 or 65 years but with no upper limitation. This sampling can result in an age range from 60 or 65 years up to 90 years plus (i.e., of 25–30 years). Even if we assume that in many of the studies on adult development and aging the sample of older adults is positively selected because they have to be sufficiently highly functioning to participate in psychology research, the differences between the subsamples of younger and older adults regarding age range as well as homogeneity concerning background variables is likely to contribute to a greater homogeneity of distributions of variables of interest in the group of younger compared with that of older adults. Moreover, the group of older adults then comprises the third and fourth age (or the "young" and "old" old). Baltes [1997] has argued and Baltes and Smith [2003] have empirically demonstrated that the psychological profiles and processes observed in these two subgroups of older adults differ significantly.

Comparing a relatively homogeneous group concerning the age span and lifestyle of younger adults with a much more heterogeneous group of older adults might contribute to Dannefer's observation that variables show more variance in older than in younger adults in cross-sectional comparisons [Dannefer, 2003]. This situation represents a statistical problem as a more restricted variance in one of the groups in the predictor and/or the criterion also produces a lower correlation coefficient in this group than in the higher-variance group. In a regression analysis, this might result in a significant age-by-variable interaction, with a lower (or no) correlation in the group of young adults (due to the restricted variance) and a significantly higher association in the group of older adults.

In fact, it can be demonstrated in a simulation that this is the case (see Appendix) and that the detection of a significant interaction of age with the variable of interest when no interaction is actually present in the data (see fig. A1 in the Appendix for the distribution of beta weights identified) occurs in more than 30% of the cases (see fig. A2 in the Appendix for the distribution of p values in the simulation). In a simulation, we can set variables to interact (or not). In contrast, in empirical studies we do not know a priori whether an interaction between age and a variable of interest is truly present or not. Thus, in order to determine whether a significant interaction reflects true association patterns in the population, one would need to have information from representative studies that provide estimates of the true distribution in the population. Unfortunately, this condition is only very rarely met. Given that the sampling of younger adults is very often less broad than the sampling of older adults, it is likely that the variance in younger adults is underestimated in many studies.

A Misguided Attempt to Investigate Developmental Processes in Age Group Comparisons: Mediational Analyses of Age-Related Differences

One common attempt to investigate potential developmental factors contributing to differences between age groups is CAVE (Cross-sectional Age Variance Extraction). In this approach, the hypothesized developmental mechanism leading to age-related differences is included as a mediator between age and an outcome variable, usually for a priori theoretical reasons. For instance, in a cross-sectional study, Zabel, Christopher, Marek, Wieth, and Carlson [2009] found that the negative cross-sectional relationship between age and financial risk taking was significantly reduced when including sensation seeking in the model. They conclude that decline in sensation seeking mediates the relationship between age and risk taking or, in other words, is one of the factors that drive developmental changes in risk taking. This hypothesis might be true, but the mediational analyses do not permit this conclusion. In line with previous work providing warnings regarding the pitfalls of this approach [Hofer & Sliwinski, 2001; Lindenberger & Pötter, 1998; Sliwinski & Hofer, 1999], Lindenberger et al. [2011] recently demonstrated formally that such mediation effects need not

DOI: 10.1159/000357177

be mapped onto longitudinal changes in outcome. The authors show mathematically that a significant mediation of cross-sectional age-related differences does not imply that the mediator variable is predictive of longitudinal changes in the outcome. Conversely, they show that even if no significant mediation effects are detected between age and the outcome variable assessed in a cross-sectional manner, the mediator might nevertheless be associated with longitudinal change in the outcome. Thus, the authors conclude that CAVE cannot be regarded as a useful approach to understanding developmental mechanisms, and that, instead, individual change over time needs to be assessed in order to circumvent the problems related to mediational analyses of cross-sectional associations.

Although we agree with this assessment, we hasten to add that longitudinal studies do not remedy all of the problems associated with a correlational approach to the study of development. Longitudinal designs cannot address whether an observed change over time is restricted to this particular cohort or generalizes over time. Moreover, multiple testing might lead to retest effects masquerading as developmental change [Hoffman, Hofer, & Sliwinski, 2011]. Moreover, so as not to miss differential rates of developmental changes depending upon the construct and the age range under study, longitudinal designs need to be strongly based on theories specifying the appropriate spacing of time intervals [Lerner et al., 2009].

A fairly new design for investigating multivariate associations on an intraindividual level (coupling of variables) in different age groups and interindividual differences in the variability of variables and their associations with other variables entails measurement bursts that comprise multiple assessments of variables over a relatively short period of time [for an early elaboration of this approach, see Nesselroade, 1991]. For instance, Brose, Schmiedek, Lövdén, and Lindenberger [2011] found in a measurement burst study over a period of 100 days that younger and older adults report more intrusive thoughts on days when they experienced stress, but that older adults suffered less negative affect from stress. Addressing the coupling of negative affect and intrusive thoughts, older adults showed a weaker association between intrusive thoughts and negative affect over the period of the study. Such results are very useful for understanding the link between daily stress, intrusive thoughts, and negative affect.

Measurement burst studies do not claim to solve the problems of cross-sectional studies. Nevertheless, it seems important to point out that, useful as they are for investigating coupling effects, measurement burst studies are also based on age-group comparisons and are correlational in nature. As such, they have similar problems as all cross-sectional studies comparing younger and older adults and they cannot provide causal information. Thus, although measurement burst designs represent a very interesting design addressing many questions of interest to psychologists studying adult development and aging, they do not solve the problems raised here and elsewhere concerning methods for studying developmental processes.

One very promising and elegant way to use the advantages of a measurement burst or daily-diary study design and to simultaneously address causal relations is to combine daily assessments with an experimental design by using within-person manipulations. Zarit, Kim, Femia, Almeida, Savla, and Molenaar [2011] have used this approach using a within-person withdrawal design (A-B-A-B) comparing days when caregivers of individuals with dementia did (A) or did not (B) make use of adult day services. They found that on A days (when individuals with dementia used adult day services) caregivers were less stressed and reported fewer behavioral and sleeping problems than on B days. The experimental portion of this design allows the causal interpretation that the use of adult day services by individuals with dementia unburdens their caregivers.

Recommendations

So what can researchers of adult development and aging – or other individual differences – do? Our major suggestion to address the problems outlined in the first section of this paper is to abandon extreme-group comparisons and include the full range of variance in a variable of interest such as age. Even if these studies are still correlational in nature, they will be less susceptible to inflated group differences due to extreme-age designs. They also will provide valuable information as to whether developmental trajectories and associations follow a more linear or nonlinear function.

Importantly, problems of inferring causality which are inherent in correlational studies need to be addressed by using experimental designs targeting the proposed psychological mechanisms underlying age-related differences. However, not all laboratory studies are created equal in terms of their potential for testing causal mechanisms related to age-related differences. In particular, the utility of such experiments may be related to their external validity. This point concerns the selection of stimulus materials for experiments and the assessment of both independent and dependent or outcome variables as well as the selection of study participants. Next, we consider the attributes that would be needed for a more fully multimethodological approach which would be well and explicitly grounded in theory for more systematic testing of theories of individual differences.

Experimental Designs Testing Age-Related Differences

There is an overwhelming consensus among researchers that the best way of testing causal relations is to manipulate the factor or process hypothesized to cause a certain outcome by using randomized control group designs; this is also true for processes driving age-related differences – yet, even though the strongest of the designs available to psychology, it is underused in this field. As pointed out above, one of the reasons may be that chronological age cannot be manipulated. However, age represents only carrier or proxy variables for psychological processes associated with chronological age [e.g., Wohlwill, 1970]. Even though chronological age cannot be manipulated, many psychological processes theorized to be associated with age do, in fact, lend themselves to experimental manipulation.

A prominent example of a research program using experimental manipulation of age-related processes is a study conducted in the context of socioemotional selectivity theory by Carstensen and colleagues [e.g., Carstensen, Isaacowitz, & Charles, 1999]. This theory posits that age-related differences in future time perspective regulate social selectivity. Complementing the cross-sectional correlational research on age-related differences in social selectivity, experimental studies manipulate future time perception and assess subsequent social selectivity [e.g., Fredrickson & Carstensen, 1990; Fung & Carstensen, 2003; Fung, Carstensen, & Lutz, 1999]. In these studies, the limitation of future time perspective simulates old adulthood in younger adults, and the extension of future time perspective young adulthood in older adults. This manipulation leads to a reversal of effects on social selectivity and thus presents an elegant demonstration of causation. A decreasing future time perspective is but one example of a developmental process that is hypothesized to drive age-related differences.

Another example for an experimental approach to adult development and aging is experimentally manipulating conditions that should compensate for the hypothesized developmental process underlying observed age differences. Research by Hess represents this approach. Hess [2005] argues that decline in fluid cognitive functions across adulthood [e.g., Li, Lindenberger, Hommel, Aschersleben, Prinz, & Baltes, 2004] causes a stronger reliance on existing cognitive schemata in information processing in older adults, which saves cognitive resources. One possible test of this hypothesis is to manipulate context conditions that increase the motivation to spend cognitive resources on the task at hand, such as public accountability or personal relevance. As predicted, experimental manipulation of the motivation to perform well decreased or eliminated schema-driven information processing [e.g., Chen, 2004; Hess et al., 2001]. To put it more generally, if it is theoretically proposed that the increasing limitation of resources associated with age underlies age-related differences, these resources can either be directly manipulated (e.g., by decreasing people's sensory functioning, as was done by Lindenberger et al. [2001]; by increasing the difficulty of a task for younger adults so that performances match the difficulty level for older adults, as was done in a study by Li, Lindenberger, Freund, and Baltes [2001]) or indirectly manipulated by creating experimental conditions that compensate for these resource limitations (as in the case of Hess' research mentioned above).

Training Studies

Training studies are special types of experiments. If the explanation for age-related differences concerns the lack of experience or, conversely, the larger amount of experience in a task in older adults, the experimental manipulation could be training studies. If older adults are believed to lack experience in a task such as certain cognitive tasks or tests, they could be trained over a specific period of time. A pre-post test comparison design involving a nontraining control group could then provide information about the effect of experience in a task on explaining age-related differences in pretraining levels.

Park and Reuter-Lorenz [2009] present an excellent example of the use of experimental and training studies in their review paper on evidence for the scaffolding theory of aging and cognition. This theory posits that neural structural decline such as decreased integrity of white matter pathways contributes to age-related differences in the functional recruitment of prefrontal brain regions for certain cognitive tasks. For obvious reasons, it would be unacceptable to disrupt white matter integrity to test this hypothesis experimentally. However, Park and Reuter-Lorenz [2009] review several experimental studies (using repetitive transcranial magnetic stimulation) that cause transient disruptions in neural functioning in support of some aspects of their theory and training studies thought to augment neural functioning for other aspects.

Although training studies have been used primarily in the cognitive area, they are, in principle, equally applicable to other research fields such as emotional or motivational development.

Taking a Brunswikian Approach

The difficulty of experimentally manipulating mechanisms driving age-related differences also points to an important, although distinct, problem in experimental studies in general, and in training or intervention studies in particular – namely, the question of the external validity of the experiment. More specifically, *external validity* concerns the question of whether an experiment is a good representation of the organism-environment relation [Brunswik, 1952]. In other words: Is the sampling of stimuli used in the experiment representative of the population of stimuli that a person encounters in his or her natural environment? According to Brunswik, this representation is crucial because psychological processes represent adaptations to stimuli in the natural environment of a person and can only be tested if the stimuli are representative of this environment. We maintain that, although proposed more than 60 years ago, this approach has not been fully recognized with regard to its usefulness to further theoretical advances in the field of adult development and aging. For this reason, we elaborate more on Brunswik's [1952] approach to experimentation in the following. We also want to point out that this approach is very much in line with the perspective of relational developmental system models [Lerner, 2012; Overton, 2013] that stresses the importance of (multidirectional) person-context transaction over time. This perspective acknowledges that the developmental context is essential for understanding how individuals develop.

Using the terminology of Brunswik's [1952] lens model (fig. 1), the relation between distal and proximal cues concerns ecological validity (i.e., the relation between an inferred state such as "emotional intelligence" and a cue like "recognition of a specific facial expression as smiling").[1] The relation between proximal cues and a person's response concerns cue utilization validity (i.e., the relation between the cue of recognizing a specific facial expression as smiling and categorizing the stimulus as happy). Both relations are probabilistic and specific to the environment of a person. For instance, for younger adults, the proximal cue of reporting that they enjoy dancing on tables at a party might be a valid indicator of extraversion. However, this interpretation might not be true for older adults, for whom this might be an indicator of nonnormative behavior. Similarly concerning cue utilization validity, the interpretation of a smiling face as indicating happiness might be highly valid in the environment of younger adults. In the environment of older adults, however, a smiling face might warrant the attribution of friendliness to the smiling person.

Central to Brunswik's approach to experimentation is the notion of *representative design*. Based on the rule that "one may generalize the results of observations only

[1] As lamented by Hammond [1998a, b] in the experimental literature, the notion of ecological validity is often used to mean that an experimental setup resemble that of the "real world" and, by virtue of this, generalize to settings outside the laboratory. Note, however, that Brunswik [1955] defined "ecological validity" as the probabilistic relation between a cue and a distal variable, thus referring to the property of stimuli. External validity denotes the generalizability of relations found in an experiment to "real-life" situations.

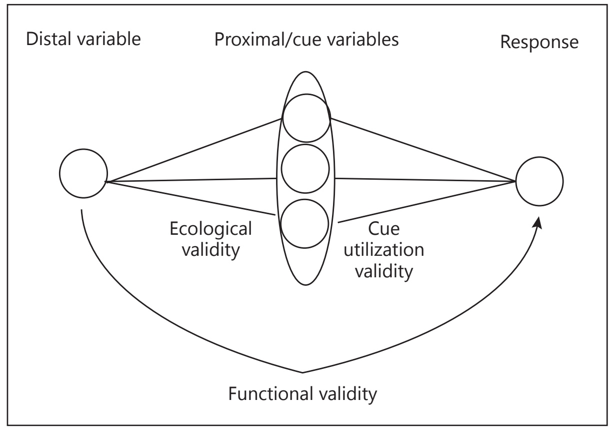

Fig. 1. Simplified depiction of Brunswik's lens model.

to those circumstances or objects that have been sampled" [Hammond, 1998a, p. 1], Brunswik proposed to sample stimuli representatively, so as to reflect the ecology of the persons under investigation regarding the frequency, range, value, distribution, and covariation of its features. Brunswik maintained that the use of stimuli which are not representative of a person's environment might alter the processes under investigation and, thereby, create misleading results. Again, to complicate matters for the study of adult development and aging, a design might well be representative for one age group but not for another, leading to age-related differences that might not be present in the natural environment.

According to Brunswik, representativeness of the stimuli can be achieved by defining a reference class of relevant stimuli as they are represented in the natural environment of the organism and then randomly drawing a sample from them so as to ensure the distribution of the stimuli or, in the case of multiple stimuli (or stimulus dimensions), their co-occurrence is also sampled. A reference class is defined as the class of objects or events that the phenomenon of interest entails. Defining the reference class is essential, since it determines which cues might be associated with the distal variable as well as their cue validity (fig. 1). For instance, when interested in age-related differences in accuracy of emotion recognition, the reference class might be the range of facial expressions and their pairing with vocal and postural emotional expressions people encounter in their everyday lives [Isaacowitz & Stanley, 2011]. Note that the likelihood of people of different ages encountering these emotional expressions might be different. In the self-selected social contexts of older adults, there might be a lower likelihood of encountering strongly negative emotional expressions encompassing facial, vocal, and postural signals than in the less self-determined educational and work contexts of younger and middle-aged adults. Thus, there might be age-related differences in the naturally occurring frequency distribution of different emotional stimuli and their co-occurrence for different age groups. This implies that when studying age-related differences, researchers need to determine reference classes for the different age groups under investigation in order to be able to sample representatively from this reference class. We are not aware of work in the area of adult development and aging that has used this approach [as one example of how this *can*

Human Development 2013;56:351–371
DOI: 10.1159/000357177

be done in psychological research in principle, see Gigerenzer, Hoffrage, & Kleinbölting, 1991, in their theory of confidence based on probabilistic mental models].

Another way of representative sampling according to Brunswik is to sample stimuli (or situations) by probing the naturally occurring, relevant situations randomly. This can be done, for instance, through time sampling procedures. This procedure consists of randomly probing participants throughout their everyday lives and assessing the variables of interest. For instance, in study 2 reported by Riediger and Freund [2008], young, middle-aged, and older adults were prompted 6 times a day on a total of 9 days (6 weekdays and 3 weekend days) to report two types of conflict of their current activities with their personal goals ideographically assessed before (i.e., "*ought to* do something else" and "*would like to* do something else"). This procedure allowed the assessment of how often these two types of conflict of activities with personal goals occur in the everyday lives of differently aged adults in their natural environments.

Hammond [1998b, p. 4] stresses the importance of a theory-based approach to the selection of stimuli: "In short, you have to be specific about the variables that a theory tells you are the important ones, namely, the ones that if ignored would produce a critically different result than if not ignored." Hammond's approach to representative sampling involves identifying the *theoretically relevant structural elements* of the stimuli and then mirroring them in the stimulus material. One example stems from recent work on age differences in sensitivity to Duchenne vs. non-Duchenne (or enjoyment vs. non-enjoyment) smiles [Murphy, Lehrfeld, & Isaacowitz, 2010]. Most studies of aging and emotion recognition focus on still images and ask participants to select from a list of emotion names to identify the emotion purportedly expressed by the face. However, this is not what people do most of the time in their everyday lives. This is an important point, because potential age-related benefits of an experience of recognizing emotions in everyday life may only be conferred to tasks that match the actual experience; it might theoretically be interesting when older adults do worse on tasks that do not tap into their experiences, but this is a different question than whether younger and older adults differ in how well they can recognize emotional stimuli such as a Duchenne and a non-Duchenne smile. This latter question can only be answered if the experimental task represents the type of judgment and the type of stimulus that are theoretically relevant to the phenomenon under investigation. In the case of emotion recognition, there are good theoretical reasons for assuming that emotions are typically seen as evolving over time. Thus, using dynamic videos of smiles and asking participants whether or not the smile expressed is genuine represents an example of a conceptually derived task for investigating a type of emotion recognition that could benefit from experience over time and, hence, might be related to age in a very different way from still pictures, which are only rarely encountered in everyday life. Indeed, the study by Murphy et al. [2010] shows that the pattern of age effects is different (and more positive for older adults) when compared with the standard static emotion recognition task.

If the stimulus material does not attain representativeness, generalizations of psychological processes to situations outside the experimental setup are not possible. While concerns about ecological validity (i.e., the relation between distal variable and proximal or cue variables; fig. 1) complicate experimentation in the study of age-related differences, we nonetheless believe that these kinds of experiment considering ecological validity are essential to further the field and make theoretical progress.

Similar arguments regarding the lack of ecological validity have been put forth by Kunzmann and her colleagues in the context of emotion recognition, empathy, and emotion regulation [Kunzmann & Grühn, 2005; Richter, Dietzel, & Kunzmann, 2010; Richter & Kunzmann, 2011], and have recently been elaborated within a Brunswikian framework by Isaacowitz and Stanley [2011] in the context of emotion recognition and aging. The basic argument is that the way emotion recognition by older adults is studied in psychological research does not match well the ways in which emotion recognition is actually done by older adults in everyday settings, because typical research tasks deprive the perceiver of a number of contextual cues (temporal, interpersonal, environmental) that offer additional sources of information to a perceiver outside the laboratory. Thus, typical research paradigms used to study age-related differences in emotion regulation and recognition may underestimate the abilities of older adults because these paradigms poorly match the way the researched process actually occurs outside the laboratory.

The same argument holds for all kinds of experiments involving group-related differences. This application is particularly obvious when studying persons with different cultural backgrounds for whom stimuli might have very different meanings. For instance, a picture of a dog might represent a negative emotional stimulus inducing disgust in the Sunni Islam culture, where dogs are seen as unclean, but a positive emotional stimulus in many Western cultures that cherish dogs as pets enriching families' lives. Thus, if different reactions to pictures of dogs are found between Sunnis and Western Christians (e.g., avoidance vs. approach behavior), the interpretation of this finding obviously needs to include the different valence of dogs in both groups. Similarly, if older adults react less strongly to certain emotional pictures, this reaction does not necessarily reflect reduced emotional reactivity but might be a function of the different meaning of emotional stimuli for different age groups. For instance, erotic pictures depicting young lovers likely have an arousing, positive valence for younger adults but might take on a nostalgic and sad quality for older adults, who might be reminded of past times that are lost forever. Thus, a study might need to employ different stimuli for different age groups. On the basis of this rationale, Kunzmann and Grühn [2005] selected different film clips of topics particularly relevant to younger or older adults and showed that age-appropriate stimuli evince the same strength of emotional reactions in younger and older adults.

Although the ecological validity and equivalence of stimuli across groups of participants is a general problem in experimental psychology [Fiedler, 2011; for cognitive psychology, see Dhami, Hertwig, & Hoffrage, 2004], the problem is even more difficult to tackle in the study of age-related differences, because what constitutes a representative design is not necessarily the same for the different age groups. As a first step, then, experiments have to establish an equivalence of representativeness of the stimuli that are used for the different age groups under investigation; this necessity does not only concern measurement equivalence across different age groups, as is aimed at primarily in the fields of psychometric cognitive and personality research [e.g., Hertzog & Nesselroade, 2003; Zimprich, Allemand, & Lachman, 2012].

Only rarely is the selection of measures and samples theoretically based, and even less often are the materials and participants sampled representatively. One of the very promising nonexperimental possibilities for capturing psychological phenomena in people's everyday lives is the use of experience sampling methods. Such methods sample experiences throughout the day as they occur in a person's natural environment. Note, however, that the external validity of the measures used in these multiple assessments are subject to the same problems as one-time assessments.

The Brunswikian approach does not include our recommendation above on experimentally manipulating hypothesized developmental mechanisms in laboratory studies, but we believe that a modified Brunswikian approach which adds this component will be especially useful to advance the field of adult development conceptually as well as methodologically.

Nonetheless, one challenge of experimental designs that may be especially important in the study of age-related differences concerns an appropriate design for testing hypothesized *processes* underlying these differences – in other words, how the Brunswikian approach is combined with the mechanistic approach. Experiments in the field of adult development and aging run the risk of leaving out important steps in the causal chain or operationalizing something other than the construct used to frame their hypotheses. This situation can lead to inappropriate conclusions. For example, studies that purport to be about certain developmental mechanisms may infer those very mechanisms from the outcomes they are proposed to cause. The flawed logic of such studies goes like this: An experiment manipulating X results in age differences in assessed outcome Y; Y is proposed to be caused by (the not-assessed) mechanism Z; age differences in outcome behavior Y are then interpreted as being *caused* by Z. However, in the absence of assessing the hypothesized mechanism Z, such an interpretation is not warranted.

One example of this issue stems from the recent growth in published papers on the age-related positivity effect in attention and memory. The vast majority of these studies report age differences in some attention or memory task, indicating that older adults attend less to (or remember fewer) negative stimuli or attend more to (or remember more) positive stimuli than younger adults. This pattern is expected for age-related positivity effects, and on a descriptive level, the design for testing for the presence of the positivity effect is appropriate. Concerns emerge when such age differences in attention or memory are framed as testing hypotheses concerning, and providing evidence for, age differences in emotion-regulatory processes [see Isaacowitz & Blanchard-Fields, 2012, for a more in-depth discussion of this issue], despite a lack of any direct measurement of emotion-regulatory processes or outcomes. Using the schema above, in these cases manipulation of the valence of emotional stimuli (variable X) leads to age differences in memory for positive vs. negative stimuli (outcome Y). Memory for positive and negative stimuli (outcome Y), in turn, is proposed to be caused by (the not-assessed) motivation to regulate emotions (process Z). Although there may be good conceptual reasons for viewing age differences in attention or memory as a product of age-related differences in emotion regulation goals, this hypothesis needs to be tested directly (e.g., by inducing emotion regulation goals) and cannot be inferred from cognitive or attentional outcomes. Similarly, emotion regulation outcomes such as improvement in positive affect or decreased negative affect cannot be inferred from cognitive processes such as attention and information processing.

Taking a Brunswikian perspective, one limitation of the literature to date on age-related positivity effects in attention and memory has been its reliance on a fairly limited number of emotional stimuli (International Affective Picture System, emotional pictures) and a narrow range of paradigms. Reliance on these stimuli raises the question of whether the paradigms sample the range of scenarios in which older individuals may (or may not) show positivity effects in their everyday lives. Evidence suggesting that the nature of possible positivity effects may vary to some degree with the choice of methodology [e.g., Depping & Freund, in press; Rovenpor, Skogsberg, &

DOI: 10.1159/000357177

Isaacowitz, 2013] heightens this concern. To take the Brunswikian approach to this topic, then, requires a broader consideration of contexts in which positivity effects can be examined, and the pattern of when and where they emerge can be used to feed back to the theory.

Experimental work on individual differences must be careful to match methods closely to hypothesized conceptual processes. If, as we suggest, proper experimentation is one of the best tools for identifying mechanisms driving individual differences, then it is critical that these experiments actually measure the mechanisms and outcomes they purport to study.

Conclusions

The study of adult development and aging presents formidable methodological challenges that go beyond the challenges faced by general psychology. A lot of research in this field uses (extreme-) group comparisons. However, given the various problems associated with this design, we urge the field (a) to use continuous sampling across the full scope of variation in the age range of interest, and (b) to complement correlational designs with experiments. Experiments can be used as a way to simulate group differences, thereby testing factors that might *cause* individual differences. In line with the recent theoretical developments in relational developmental systems models, we join Fiedler [2011] and Isaacowitz and Stanley [2011] in their plea for using a more Brunswikian approach to experiments in order to avoid the pitfalls associated with the use of nonrepresentative experimental stimuli and manipulations.

We urge researchers in the field of adult development and aging to make use of a Brunswikian approach to experimentation that takes into account the ecology of adults in different age groups in order to achieve a representative sampling of stimuli that is adequate for all age groups under investigation. This approach allows investigating age-related differences in the processes that may constitute specific adaptations to the respective ecologies of different age groups. Such an approach also fits with relational developmental systems models [e.g., Lerner, 2012; Overton, 2013] that have primarily been adopted as a frame in the study of development in childhood and adolescence but are less prominent in research on adult development and aging. Taking the relational developmental systems perspective, psychological theories need to specify the multidirectional transactions between a person and the various levels of his or her context. On the basis of such specifications, researchers can then operationalize and experimentally manipulate the factors believed to contribute to age-related differences, including testing the mechanisms in a range of contexts that include a formal consideration of how the person-context transaction varies throughout adulthood.

The use of an adequate methodology lies at the heart of good research. It does not only contribute to the accumulation of empirical findings describing phenomena of interest, but provides a foundation for testing theories as well. We join Greenwald [2012] in the assessment that "there is nothing so theoretical as a good method" (p. 99) and add that good methods will especially be critical to the continued development of knowledge about individual differences in general and adult development and aging in particular.

Appendix

A simulation with 1,000 runs was set up in order to test if restricting the range in one of two subsamples differing in age increases the likelihood of detecting an age-by-predictor interaction even though no interaction is present in the data. The final sample sizes in both groups were set to 100. In the younger adults, the range was restricted by generating an original sample of n = 150, and then selecting n = 100 from the original sample by omitting the 2 × 25 subjects at the two extremes of the distribution. In this simulation, the true correlation between the predictor and the criterion within each of the two age groups was set to $\rho = 0.5$. The true beta of the interaction of the predictor with age in predicting the criterion was 0.

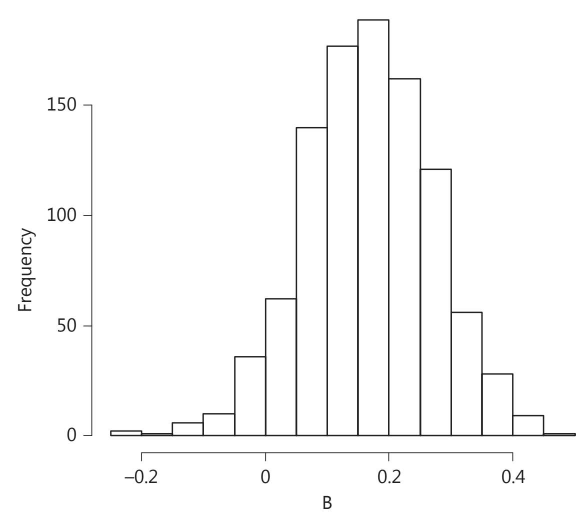

Fig. A1. Distribution of regression weights of the interaction term of the variable of interest with age obtained in the simulation of a group comparison with restricted variance.

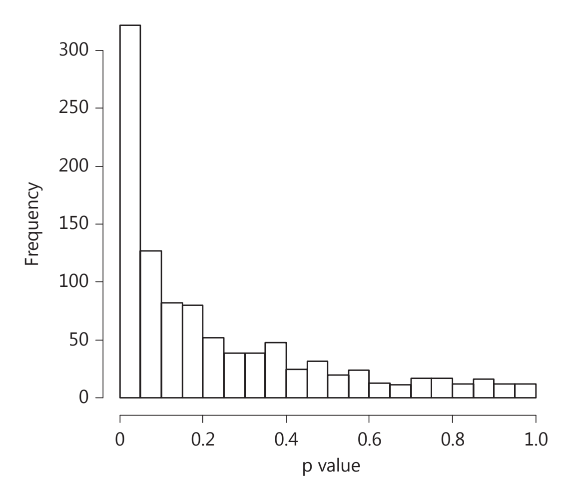

Fig. A2. Distribution of p values for the interaction term of the variable of interest with age obtained in the simulation of a group comparison with restricted variance.

DOI: 10.1159/000357177

The model was specified as:

Predictor: x(young) ~ N(mean = 0, SD = 2); x(old) ~ N(mean = 2, SD = 2)

Criterion (y) = rho · x + sqrt(1 – ρ) · N(0, 2)

Young: sample restricted to values of x in range [–2, 2] to generate restriction of range

Regression model: y = a + B1 · x + B2 · age + B3 · (age · x)

The p value refers to the significance of the interaction B3 (age by predictor). The histograms show the interaction terms and the associated p values across 1,000 replications of the simulation. More than 30% (322 out of 1,000) simulations yielded $p < 0.05$, even though no interaction was specified in the data.

The simulation code may be obtained by the first author upon request.

Acknowledgements

We are grateful to Klaus Oberauer for discussions of the issues presented in this paper and for running and summarizing the simulation reported in the Appendix. We also wish to thank Jana Nikitin, Martin J. Tomasik, and David Weiss for their comments on an earlier draft of this paper.

References

Baltes, P.B. (1968). Longitudinal and cross-sectional sequences in the study of age and generation effects. *Human Development, 11,* 145–171.

Baltes, P.B. (1997). On the incomplete architecture of human ontogeny: Selection, optimization, and compensation as foundation of developmental theory. *American Psychologist, 52,* 366–380.

Baltes, P.B., Reese, H.W., & Lipsitt, L.P. (1980). Life-span developmental psychology. *Annual Review of Psychology, 31,* 65–110.

Baltes, P.B., Reese, H.W., & Nesselroade, J.R. (1977). *Life-span developmental psychology: An introduction to research methods*. Monterey: Brooks Cole.

Baltes, P.B., & Smith, J. (2003). New frontiers in the future of aging: From successful aging of the young old to the dilemmas of the fourth age. *Gerontology, 49,* 123–135.

Birren, J.E. (1999). Theories of aging: A personal perspective. In V.L. Bengtson & K.W. Schaie (Eds.), *Handbook of the psychology of aging* (3rd ed., pp. 3–20). San Diego: Academic Press.

Brose, A., Schmiedek, F., Lövdén, M., & Lindenberger, U. (2011). Normal aging dampens the link between intrusive thoughts and negative affect in reaction to daily stressors. *Psychology and Aging, 26,* 488–502.

Brunswik, E. (1952). The conceptual framework of psychology. *International encyclopedia of unified science. Vol. 1, No. 10* (pp. 656–760). Chicago: University of Chicago Press.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review, 62,* 193–217.

Carstensen, L.L., Isaacowitz, D.M., & Charles, S.T. (1999). Taking time seriously: A theory of socioemotional selectivity. *American Psychologist, 54,* 165–181.

Chen, Y. (2004). Age differences in the correction of social judgments: Source monitoring and timing of accountability. *Aging, Neuropsychology, and Cognition, 11,* 58–67.

Dannefer, D. (2003). Cumulative advantage/disadvantage and the life course: Cross-fertilizing age and social science theory. *Journals of Gerontology: Series B, Psychological Sciences and Social Sciences, 58,* S327–S337.

Depping, M., & Freund, A.M. (in press). When choice matters: Task-dependent memory effects in older adulthood. *Psychology and Aging*.

Dhami, M.K., Hertwig, R., & Hoffrage, U. (2004). The role of representative design in an ecological approach to cognition. *Psychological Bulletin, 130,* 959–988.

Elashoff, J.D. (1969). Analysis of covariance: A delicate instrument. *American Educational Research Journal, 6,* 383–401.

Fiedler, K. (2011). Voodoo correlations are everywhere – not just in neuroscience. *Perspectives on Psychological Science, 6,* 163–171.

Fredrickson, B.L., & Carstensen, L.L. (1990). Choosing social partners: How old age and anticipated endings make people more selective. *Psychology and Aging, 5,* 335–347.

DOI: 10.1159/000357177

Freund, A.M., & Ritter, J.O. (2009). Midlife crisis: A debate. *Gerontology, 55,* 582–591.
Fung, H.H., & Carstensen, L.L. (2003). Sending memorable messages to the old: Age differences in preferences and memory for advertisements. *Journal of Personality and Social Psychology, 85,* 163–178.
Fung, H.H., Carstensen, L.L., & Lutz, A.M. (1999). Influence of time on social preferences: Implications for life-span development. *Psychology and Aging, 14,* 595–604.
Gerstorf, D., Ram, N., Hoppmann, C., Willis, S.L., & Schaie, K.W. (2011). Cohort differences in cognitive aging and terminal decline in the Seattle Longitudinal Study. *Developmental Psychology, 47,* 1026–1041.
Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review, 98,* 506–528.
Greenwald, A.G. (2012). There is nothing so theoretical as a good method. *Perspectives on Psychological Science, 7,* 99–108.
Hammond, K.R. (1998a). *Ecological validity: Then and now.* Retrieved from http://www.brunswik.org/notes/essay2.html.
Hammond, K.R. (1998b). *Representative design.* Retrieved from http://www.brunswik.org/notes/essay3.html.
Hertzog, C., & Nesselroade, J.R. (2003). Assessing psychological change in adulthood: An overview of methodological issues. *Psychology and Aging, 18,* 639–657.
Hess, T.M. (2005). Memory and aging in context. *Psychological Bulletin, 13,* 383–406.
Hess, T.M., Rosenberg, D.C., & Waters, S.J. (2001). Motivation and representational processes in adulthood: The effects of social accountability and information relevance. *Psychology and Aging, 16,* 629–642.
Hofer, S.M., & Sliwinski, M.J. (2001). Understanding ageing: An evaluation of research designs for assessing the interdependence of ageing-related changes. *Gerontology, 47*, 341–352.
Hoffman, L., Hofer, S.M., & Sliwinski, M.J. (2011). On the confounds among retest gains and age-cohort differences in the estimation of within-person change in longitudinal studies: A simulation study. *Psychology and Aging, 26,* 778–791.
Isaacowitz, D.M., & Blanchard-Fields, F. (2012). Linking process and outcome in the study of emotion and aging. *Perspectives on Psychological Science, 7,* 3–17.
Isaacowitz, D.M., & Stanley, J.T. (2011). Bringing an ecological perspective to the study of aging and emotion recognition: Past, current, and future methods. *Journal of Nonverbal Behavior, 35,* 261–278.
Jackson, J.J., Thoemmes, F., Jonkmann, K., Lüdtke, O., & Trautwein, U. (2012). Military training and personality trait development: Does the military make the man, or does the man make the military? *Psychological Sciences, 23,* 270–277.
Kunzmann, U., & Grühn, D. (2005). Age differences in emotional reactivity: The sample case of sadness. *Psychology and Aging, 20,* 47–59.
Lerner, R.M. (2012). Developmental science: Past, present and future. *International Journal of Developmental Science, 6,* 29–36.
Lerner R.M., Schwartz, S.J., & Phelps, E. (2009). Problematics of time and timing in the longitudinal study of human development: Theoretical and methodological issues. *Human Development, 52,* 44–68.
Li, K.Z., Lindenberger, U., Freund, A.M., & Baltes, P.B. (2001). Walking while memorizing: Age-related differences in compensatory behavior. *Psychological Science, 12,* 230–237.
Li, S.C., Lindenberger, U., Hommel, B., Aschersleben, G., Prinz, W., & Baltes, P.B. (2004). Transformations in the couplings among intellectual abilities and constituent cognitive processes across the life span. *Psychological Science, 15,* 155–163.
Lindenberger, U., & Pötter, U. (1998). The complex nature of unique and shared effects in hierarchical linear regression: Implications for developmental psychology. *Psychological Methods, 3,* 218–230.
Lindenberger, U., Scherer, H., & Baltes, P.B. (2001). The strong connection between sensory and cognitive performance in old age: Not due to sensory acuity reductions operating during cognitive assessment. *Psychology and Aging, 16,* 196–205.
Lindenberger, U., von Oertzen, T., Ghisletta, P., & Hertzog, C. (2011). Cross-sectional age variance extraction: What's change got to do with it? *Psychology and Aging, 26,* 34–47.
McArdle, J.J. (2008). Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology, 60,* 577–605.
Miller, G.A., & Chapman, J.P. (2001). Misunderstanding analysis of covariance. *Journal of Abnormal Psychology, 110,* 40–48.
Murphy, N.A., Lehrfeld, J., & Isaacowitz, D.M. (2010). Recognition of posed and spontaneous dynamic smiles in young and older adults. *Psychology and Aging, 25,* 811–821.
Nesselroade, J.R. (1991). The warp and woof of the developmental fabric. In R. Downs, L. Liben, & D.S. Palermo (Eds.), *Visions of aesthetics, the environment, and development: The legacy of Joachim F. Wohlwill* (pp. 213–240). Hillsdale: Erlbaum.
Overton, W.F. (2013). A new paradigm for developmental science: Relationism and relational-developmental systems. *Applied Developmental Science, 17,* 94–107.

Park, D.C., & Reuter-Lorenz, P. (2009). The adaptive brain: Aging and neurocognitive scaffolding. *Annual Review of Psychology, 60,* 173–196.
Preacher, K.J., Rucker, D.D., MacCallum, R.C., & Nicewander, W.A. (2005). Use of the extreme groups approach: A critical reexamination and new recommendations. *Psychological Methods, 10,* 178–192.
Richter, D., Dietzel, C., & Kunzmann, U. (2010). Age differences in emotion recognition: The task matters. *Journals of Gerontology: Series B, Psychological Sciences and Social Sciences, 66,* 48–55.
Richter, D., & Kunzmann, U. (2011). Age differences in three facets of empathy: Performance-based evidence. *Psychology and Aging, 26,* 60–70.
Riediger, M., & Freund, A.M. (2008). Me against myself: Motivational conflicts and emotional development in adulthood. *Psychology and Aging, 23,* 479–494.
Rovenpor, D., Skogsberg, N., & Isaacowitz, D.M. (2013). The choices we make: An examination of situation selection in younger and older adults. *Psychology and Aging, 28,* 365–376.
Schaie, K.W. (1965). A general model for the study of developmental problems. *Psychological Bulletin, 64,* 92–107.
Schaie, K.W., & Baltes, P.B. (1975). On sequential strategies in developmental research: Description or explanation? *Human Development, 18,* 384–390.
Simmons, J.P., Nelson, L.D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22,* 1359–1366.
Sliwinski, M.J., & Hofer, S.M. (1999). How strong is the evidence for mediational hypotheses of age-related memory loss? *Gerontology, 45,* 351–354.
Stuart, E.A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science, 25,* 1–21.
Wilson, R.S., Beck, T.L., Bienias, J.L., & Bennett, D.A. (2007). Terminal cognitive decline: Accelerated loss of cognition in the last years of life. *Psychosomatic Medicine, 69,* 131–137.
Wohlwill, J.F. (1970). The age variable in psychological research. *Psychological Review, 77,* 49–64.
Zabel, K.L., Christopher, A.N., Marek, P., Wieth, M.B., & Carlson, J.J. (2009). Mediational effects of sensation seeking on the age and financial risk-taking relationship. *Personality and Individual Differences, 47,* 917–921.
Zarit, S.H., Kim, K., Femia, E.E., Almeida, D.M., Savla, J., & Molenaar, P.C.M. (2011). Effects of adult day care on daily stress of caregivers: A within-person approach. *Journals of Gerontology: Series B, Psychological Sciences and Social Sciences, 66,* 538–546.
Zimprich, D., Allemand, M., & Lachman, M.E. (2012). Factorial structure and age-related psychometrics of the MIDUS personality adjective items across the life span. *Psychological Assessment, 24,* 173–186.

DOI: 10.1159/000357177